

Assessment of Google and Microsoft Bing Translation of Journalistic Texts

Zakaryia Mustafa Almahasees

Abstract—Machine Translation (MT) systems are commonly utilized by end users since MT is available freely or at a low cost. The increasing demand for MT services nowadays means that ensuring the acceptability of the output to the potential users of such systems is a necessary task. The paper evaluates the capacity of two prominent systems, Google Translate and Microsoft Bing Translator, in producing acceptable English translations of journalistic texts written in Arabic. To do so, the study has adopted Linguistic Error Analysis of Reference [1] and of Error Classification described in Reference [2]. The results of the study show that both systems obtain outstanding results with > 90 percentage accuracy in the area of orthography and grammar. In addition, both systems obtain good results in the areas of lexical and grammatical collocations of 79.8% for Google and 74.5% for Bing. The two systems achieved good results in these categories because they have recently adopted Neural Machine Translation, which imitates the human brain to perform translation and learns from previously translated texts by humans. For future research, the study recommends conducting more assessment on translation in a variety of fields of knowledge using Linguistic Error Analysis. Machine Translation is still far from reaching fully automatic translation of a quality obtainable by human translators.

Index Terms—Google translate, assessment, microsoft bing, Machine Translation Evaluation (MTE), machine translation errors.

I. INTRODUCTION

Reference [3] indicates that Globalization made English an international language. It is the medium of instruction at most global universities and the primary means of communication across the globe. In the case of people who do not master English, MT tends to be used as a translation aid in both studying and general reading. Machine Translation (MT) evaluation is the study of translations produced by translation systems across languages. Reference [4] defines MT as “the use of computers in the process of translation from one natural language into another.” MT allows potential users to quickly translate full documents either freely as in non-commercial MT systems, or at a lower cost than human translators. The main aim of MT is to generate a translation that is similar to human translation, which is acceptable by human translators, clients and readers. Reference [4], Product Lead of Google Translate, shows that there are 500 million

users of Google Translate a day; Microsoft Bing Translator users are at about 18 million per day. Moreover, reference [4] mentions that most translation is conducted “between English and Spanish, Arabic, Russian, Portuguese, and Indonesian.” English-Arabic translation being mentioned, research on quality of MT in this language pair is therefore important.

Google provides translation for 103+ languages in different translation modes, such as speech recognition translation, sign translation and others. Microsoft Bing provides translation for 60+ languages. In the case of English and Arabic, reference [5] lists 56 MT systems that translate in both directions, English<>Arabic. Such a widespread availability of MT raises the essential question of whether MT systems are capable of providing acceptable translations that meet users’ expectations or not. Specifically, what are the limitations and strengths of the two systems here under study when handling journalistic texts from Arabic into English?

Historically, MTE (Machine Translation Evaluation) started in the second half of the 20th century and proposed many methods of evaluating MT results. MTE works to trace the improvement of MT in providing acceptable translations to the potential users. It has been conducted manually as well as automatically. Manual evaluation assesses the usability of MT systems via human participation in interviews and tests. Furthermore, it also analyses MT output in terms of linguistic error analysis. Automatic evaluation, in contrast, evaluates MT output in terms of text similarity metrics to find the degree of similarity of an MT output to a referenced, i.e. human, translation. However, there is no wholly agreed-upon method between different MT researchers.

Reference [3] shows that despite the fact that there is no universal method of automatic MT evaluation, evaluators nevertheless agree on mistranslations and errors in syntax or vocabulary. Therefore, there is an overlapping area of agreement. Reference [6] indicates that MTE relies on three main methods: design evaluation, development evaluation and evaluation by potential purchasers of MT. The design examines the system’s performance concerning a range of possible inputs. It is the first method to evaluate the basic design and the capacity of the system. Secondly, development evaluation tests the capacity of MT output in different contexts and environments. In reference to [6], the authors add that evaluation of the development of the system assesses the “economic viability of the system as a commercial product, potential sales, and leasing agreements.” Finally, purchasers’ evaluation highlights whether the system achieves the required expectations of the customers and traces the required improvements of the system.

This is the reason why the study at hand analyses MT output of the two chosen systems here under study to shed

Manuscript received June 25, 2018; revised September 11, 2018.

Zakaryia Mustafa Almahasees is with the University of Western Australia, School of Humanities, Research Grants (Research Travel Grant), Zakaryia Almahasees is with University of Western Australia, Australia (e-mail: zakaryia.almahasees@research.uwa.edu.au).

light on the pitfalls and the strengths of MT in providing an acceptable translation of the journalistic texts of Petra News Agency of Jordan.

II. REVIEW OF RELATED LITERATURE

Evaluation of MT is the most important task in the life cycle of any MT system. Reference [7] mentions that manual evaluation is considered subjective while automatic evaluation is considered an objective method of evaluation. However, while that research contends that automatic evaluation appears to be objective, the criteria and the studies show that automatic evaluation provides objective results through texts' similarity. Evaluation based on textual similarity is incapable of providing feedback on the capacity of MT systems because it analyses only correspondence between words, not quality in terms of language and semantic cohesion. Reference [7] mentions that manual evaluation covers error analysis, which provides consistent results. The study contends that error analysis is essential in tracing the learning development in human learners. Thus, it is also a crucial element for testing the capacity of MT systems and tracing the necessary improvements for any system to provide objective feedback about MT quality.

Much research has been conducted on English\leftrightarrowArabic MT translation based on automatic evaluation, but manual evaluation specifically through error analysis is still scarce. Reference [8] discusses the problem of English-Arabic translation of idioms and proverbs, finding that MT has serious drawbacks in this area. Reference [9] develops an English-Arabic MT system to translate political texts from English into Arabic. On the other hand, reference [10] tackles the ability of MT to translate English compounds into Arabic due to their frequent use in all text types. Reference [11] sheds light on the importance of creating a new system to program English-Arabic MT. Reference [12] develops a system to translate abstracts of texts in the Artificial Intelligence domain from English into Arabic. However, the effort to evaluate Arabic MT is still limited in comparison to Western contributions to MTE. Reference [13] evaluates the ability of three Arabic MT systems in dealing with English-Arabic translation. Reference [13] found that MT is still suffering from serious drawbacks, especially in grammar and semantic coherence of the translated sentences. Reference [14] assesses Google's capacity for the translation of legal texts. Reference [14] found that Google is still limited in providing acceptable legal translation as such practice require high accuracy and precision that Google fails to obtain. Notwithstanding, MT can help in providing the general meaning of the text, but the intended meaning is absent. Reference [3] indicated also that one of the main problems is the context and specific domain terms, which are unattainable by MT due to the cultural setting of each culture and what is acceptable in one language may be unacceptable in the other.

III. METHODOLOGY

This section presents the main steps taken to carry out this study. After selecting a corpus covering a wide range of journalistic texts, extracts have been inputted into Google

Translate and Microsoft Bing. Then, the researcher scrutinizes and evaluates the chosen extracts to find the errors committed by the two systems. Reference [2] has proposed five ways to identify errors: data collection, error identification, error description, error explanation and the evaluation of errors. However, any error analysis requires a systematic framework to facilitate the process of error categorizations. To this end, the study has adopted the framework of error analysis described in Reference [1]. The three chosen error categories are orthography, lexis and grammar. Furthermore, the study at hand classifies errors into minor and major errors. Minor errors are technical errors, but ones, which do not alter the flow of the text or inhibit its comprehension; these carry a weight of one mark. On the other hand, major errors are errors that do disrupt the flow of the text, but the text is still easy to comprehend; these carry a weight of two marks. More importantly, the flow of language components should not only present a good grammatical structure, but also convey the intended message adequately. Above all else, if the chosen system distorts the source text message and fails to render the message clearly, the indicated percentage will be zero mark for the system under study in the chosen extract.

A. Discussion

Given the above, the analysis of the study is conducted through three main categories of linguistics error analysis: orthography, lexis and grammar. Firstly, orthography is the study of writing conventions, such as spelling, capitalization, and punctuation, hyphenation and word breaks. Secondly, lexis is the study of all possible words of a language. Lexis error analysis includes omitted, added, and untranslated words, as well as lexical collocations, word choice, and mistranslation. Thirdly, grammar is the study of the language structure. It studies the word constructions in terms of subject verb agreement, verb conjugation, clauses, grammatical collocations and the like.

B. Orthography

Reference [15] defines orthography as "the way words are spelt or should be spelt." However, Reference [16] presents another definition of orthography: "the art of writing words [...] to accepted usage."

First extract:

The following example is extracted from the local section of Jordan Petra News agency; published on 16th October 2017.

ترأس رئيس الوزراء الدكتور هاني الملقي اجتماع المجلس الاعلى للدفاع المدني وبحضور وزير الداخلية غالب الزعبي ووزير الدولة لشؤون الاعلام الدكتور محمد المومني ومدير عام الدفاع المدني اللواء مصطفى البرازيه وكافة اعضاء المجلس والذي عقد في المديرية العامة للدفاع المدني صباح اليوم الاثنين لمناقشة استعدادات اجهزة الدولة لفصل الشتاء القادم
واعرب رئيس الوزراء عن الاعتراز والفخر بالمستوى الذي وصلت اليه المديرية العامة للدفاع المدني مؤكدا ان هذا الصرح الوطني شهد بفضل الرعاية الملكية السامية تطورا ملحوظا ومتميزا وجاهزية عالية وتجهيزات توازي مثيلاتها في ارقى الدول
وقال ان هذا الجهد لرفع اداء جهاز الدفاع المدني يعكس مدى الاهتمام بالمواطن وممتلكاته وخدمته وضمن سرعة التجاوب مع المتطلبات الاساسية والرئيسية للمواطنين في حالة الحوادث

Reference [17]

C. Google Translate

Prime Minister Dr. Hani al-Mulqi held a meeting of the Supreme Council of Civil Defense in the presence of Interior Minister Ghaleb Al-Zu'bi, Minister of State for Information Affairs Dr. Mohammad Al-Momani, Director General of Civil Defense Major General Mustafa Al-Bazaiha and all members of the Council. .

The Prime Minister expressed pride and pride in the level reached by the Directorate General of Civil Defense, stressing that this national edifice has witnessed thanks to the royal care of the high and remarkable development and high readiness and equipment equivalent to those in the highest countries.

He said that this effort to raise the performance of the civil defense system reflects the interest of the citizen and his property and service and ensure rapid responsiveness to the basic requirements of citizens in case of accidents.

D. Microsoft Bing

The prime Minister, Dr. Hani Al-Ali, presided over the

meeting of the Supreme Council for Civil Defence, with Interior Minister Ghalib Zoubi and the Minister of State for information, Dr. Mohammed Al-Moumani, director General of Civil Defence, Major General Mustafa al-Baza'a and all members of the Council, held at the Directorate General of Civil Defence Monday morning to discuss preparations for the State organs for the coming winter.

The Prime Minister expressed pride and pride in the level reached by the General Directorate of Civil Defence, asserting that this national edifice thanks to high-quality care, there has been a remarkable and distinct development, with high levels of readiness and equipment comparable to those in the finest states.

This effort to raise the performance of the civil defence system reflected the extent of concern for the citizen, his property and service, and the speed of responsiveness to the basic and key requirements of citizens in the event of accidents, he said.

Secondly, the orthographic errors are shown in Table I.

TABLE I: ORTHOGRAPHIC ERRORS

Errors	Capitalization	Punctuation	Spelling	Hyphen	Word Break	SUM	N. of Words	PER
MT								
Google	2	1	0	0	0	3	111	2.70%
Bing	3	1	1	0	0	4	111	3.60%

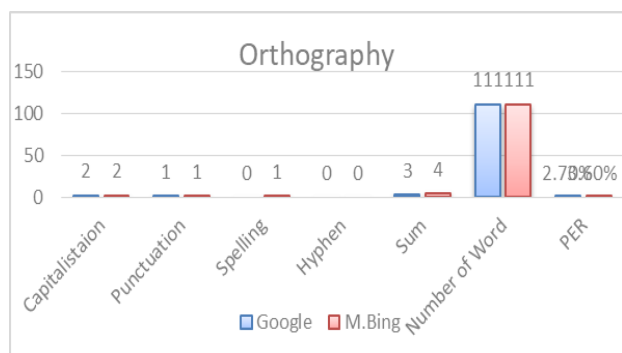


Fig. 1. Orthographic errors.

Fig. 1 illustrates the way Google and Microsoft Bing have dealt with the translation of whole extracts at the orthographic level. The study's analysis showed that Google made capitalisation errors. For instance, for the Arabic word /اداء الدفاع المدني/, Google provides an accurate translation, but it neglects rules of capitalisation; the name of organisations and institutions should be capitalised. Furthermore, Microsoft Bing also made errors in capitalisation, for example in its translation of the name of director of Jordan Civil Defence, Major General Mustafa al-Baza'a; it is clear here that Microsoft Bing does not go by the rule that states that proper names should be capitalised.

The number of errors (MN) is summed and divided by the

number of words (WN) of the whole extract. The resulting number is given as a percentage (PER %) in the following formulae:

$$\text{Google Result: } \frac{3}{111} = 2.70\%$$

$$\text{Bing Result: } \frac{4}{111} = 3.60\%$$

The result of both systems shows that orthography errors are below 4% in both systems, which indicate that Google and Microsoft Bing achieved outstanding results with >95% accuracy in rendering Arabic into English. This is a very normal amount of errors for language processing systems.

E. Lexis

Lexis is the study of the vocabulary of a language that has meaning and grammatical function. Reference [18] defines lexis as an area, which is concerned with the nature, meaning, history and use of words. Reference [19] shows that lexis is a branch of linguistics concerned with meaning and use of words. Moreover, Reference [20] indicates that lexis is "another term for vocabulary and lexis is all the words and phrases of a particular language."

Table III illustrates the errors made at the lexical stage. Lexis errors are considered major errors because lexis errors disrupt the flow of the text, and inhibits the intelligibility of the text.

TABLE II: LEXIS ERRORS

Errors	Omission	Addition	Untranslated Word	Lexical Collocations	Mistranslation	SUM	N. Words	PER %
MT Systems								
Google	2	2	0	14	6	24	111	21.6%
Microsoft Bing	2	2	0	18	6	28	111	25.2%

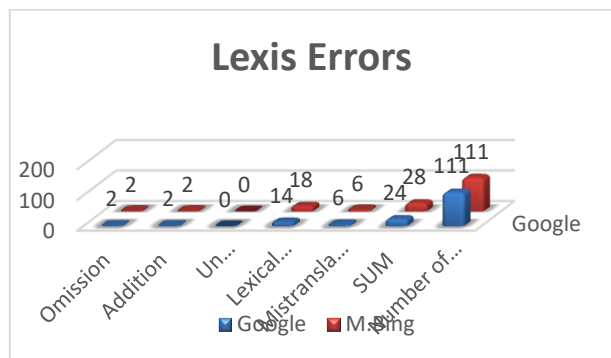


Fig. 2. Lexis errors.

Fig. 2 demonstrates a rise in a number of errors in comparison to the orthographic errors. Lexis errors include omission, addition, lexical collocations and mistranslation. Omission and addition account for the lowest percentage of lexis errors. For instance, Google translates the first part of the chosen extract, but it omits the following phrase:

صباح اليوم الاثنين لمناقشة استعدادات اجهزة الدولة لفصل الشتاء القادم

In this regard, Google omits an essential part of the meaning of the paragraph. Thus, Google translation does not convey the ST message, which inhibits the intelligibility of the text.

On the other hand, the highest percentage of lexis errors occur in lexical collocations. Lexical collocations are the ways words combine to form natural-sounding speech and writing. For example, English has the collocations “heavy meal” and “strong coffee”; it would not be normal to say “heavy coffee” or “strong meal”. In the first extract, Google committed seven errors at the level of lexical collocations. For example, Google translates the following collocation (ارقي الدول) into “the highest countries”. Google here considers (ارقي) as high-ارتفاع which indicates the violation of source text meaning: (ارقي) to be “developed. Oxford Collocations Dictionary for Students of English [20] indicates that the word “country” could collocate with “developed, underdeveloped, advanced, Third World, Communist, capitalist...” Therefore, it is unacceptable to collocate “high” with countries. Furthermore, Microsoft Bing commits the same error in rendering ارقى الدول as “the finest states”, which contradicts the meaning of the source text to wit, “developed countries”. However, “finest” cannot collocate with “state”; a state could be beautiful, but not the “finest”. The researcher recommends developed countries for ارقى الدول due to its best description of how the country industrially and economically developed. Moreover,

Microsoft Bing commits an error in word clause; the word “الحوادث” has two equivalent terms in English, “accident” and “emergency.” The context of “الحوادث” indicates “emergency” while Microsoft Bing understands it as “accident.” Despite the bad translation of the above-mentioned example, Google and Bing nevertheless achieved good results with 79.8% for Google and 74.5% for Microsoft Bing.

F. Grammar

Grammar is the study of structural conventions governing the rules of clauses, phrases and words in a language. It includes the relevant branches of language, such as phonology, morphology, syntax, semantics and pragmatics.

The Macquarie Dictionary [16] defines grammar as “the features of a language (sounds, words, formation and arrangement of words)”. The analysis of grammar in this section includes subject-verb agreement, verb clause, relative clause, preposition, word clause and grammatical collocations. Table III lists grammar errors found in the first extract of the journalistic texts.

TABLE III: GRAMMATICAL ERRORS

Verb	Relative Clause	Preposition	Grammatical Collocations	Word Clause	Number of Words	Per
4	0	0	0	0	111	3.60%
4	0	0	0	2	111	3.60%

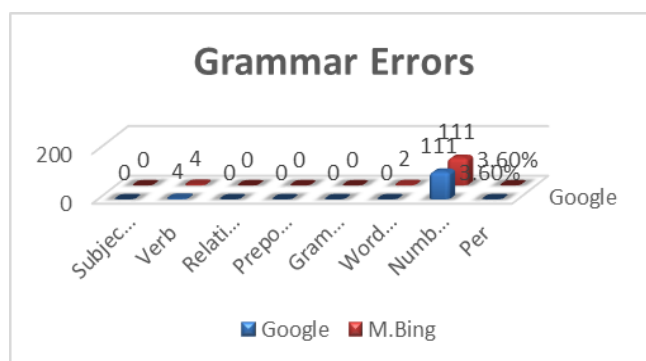


Fig. 3. Grammar errors.

Fig. 3 demonstrates the capacity of the two chosen systems to handle whole text translation. The results show that both systems achieved good results in providing a translation with reference to English rules of grammar. In this regard, it is worthy mentioned that one of the early approaches used to advance MT at the early days is Rule Based Machine Translation (RBMT) designed based on linguistic dictionaries and corpora that cover the main syntactic, semantic and morphological regularities of the source and text language respectively. The hence indicates that that grammar by Neural Based Machine Translation, the advanced MT approach, will be handled well.

In Grammar, pronouns a word that substitutes for a noun or noun, and they are classified by person, number, gender and case. The following phrase واعرب رئيس الوزراء عن الاعتزاز بنوال الفخر is taken from the above extract. Google and Microsoft Bing committed the same error in omitting the possessive adjective pronoun (his) to substitute the prime minster. Moreover, Google commits errors in adding the definite article, “the”, to Prime Minister. The definite article here is unnecessary because the context here does not describe or explain the role of the prime minster; “Prime Minster” is an official title, so there is no need for the article “the”. This shows that MT still has drawbacks in contextual translation. However this is an advanced issue of grammar; the use of the definite article in English being one of the most complex. The overall assessment of the capacity of MT to translate various extracts is shown in Fig. 4 & 5.

The above chart demonstrates that both systems achieved outstanding results in the area of orthography, and fair results at lexis and grammar levels. However, the existence of major errors inhibited the flow of the text, perhaps because the

systems might not have processed such words before. Errors might occur in specific cultural and contextual words, which are difficult for MT to deal with effectively. On the other hand, the two systems achieved good results due to the use of Neural Machine Translation, which aims to build large neural network that predicts the probability of sequence of words to read a sentence and have a correct translation as an output.

IV. CONCLUSION

The results of the study showed that MT achieved outstanding results of 92% to 93% for both systems at the orthography and Grammar levels. However, there are major linguistic errors that inhibit the comprehension of the text, such as lexical and grammatical collocations. Thus, both systems achieve outstanding results at the word level, but only good results at collocation units and cultural and specific translation levels with 79.8% for Google and 74.5% for Bing. It is worth mentioning that the results are limited for a small size of data as a part of future work planning on collecting more extracts to verify the capacity of MT in Journalistic texts. Furthermore, the study also recommends collaboration between companies producing MT technologies and translation scholars.

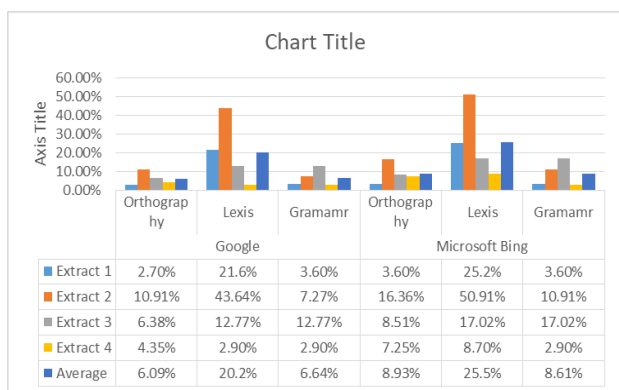


Fig. 4. Overall assessment of errors made by Google and Microsoft Bing.

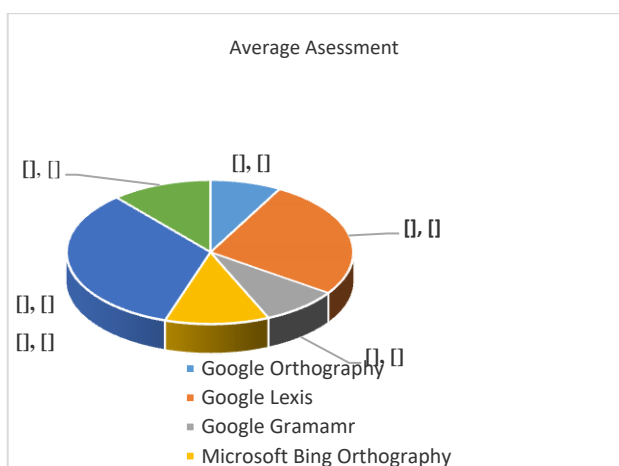


Fig. 5. Average assessment.

ACKNOWLEDGMENT

My profoundest gratitude is due to my Supervisor Prof. Dr. Helene Jaccocard for her extremely intellectual and generosity, which help me to broaden my understanding of

MT field thematically and systematically.

REFERENCES

- [1] Â. Costa, W. Ling, T. Lu í, R. Correia, and L. Coheur, "A linguistically motivated taxonomy for machine translation error analysis," *Machine Translation*, vol. 29, no. 2, pp. 127-161, 2015.
- [2] C. Dulay and K. Burt, "Error and strategies in child second language acquisition," *TESOL Quarterly*, vol. 8, pp. 129-138, 1974.
- [3] Z. Almahasees and Zakaryia Mustafa, *Machine Translation Quality of Khalil Gibran's the Prophet*, 2017.
- [4] A. Zakaryia, "Assessing the translation of Google and Microsoft Bing in translating political texts from Arabic into English," *International Journal of Languages, Literature and Linguistics*, vol. 3, no. 1, pp. 1-4, 2017.
- [5] T. Barak. (2016). 11 Google translate facts you should know. [Online]. Available: <http://www.k-international.com/blog/google-translate-facts>
- [6] J. Hutchins. (2010). Compendium of translation software. [Online]. Available: <http://www.hutchinsweb.me.uk/Compendium.htm>
- [7] W. J. Hutchins and H. L. Somers, "An introduction to machine translation," vol. 362, London: Academic Press, 1992.
- [8] N. Mostofian, *A Study on Manual and Automatic Evaluation Procedures and Production of Automatic Post-editing Rules for Persian Machine Translation*, 2017.
- [9] M. Ibrahim, "A fast and expert machine translation system involving Arabic language," Ph. D. Thesis, United Kingdom: Cranfield Institute of Technology, 1991.
- [10] A. Rafea, M. Sabry, R. El-Ansary, and S. Samir, "A machine translator for middle east news," in *Proc. the 3rd International Conference 93 and Exhibition on Multi-lingual Computing, the Centre for Middle Eastern and Islamic Studies*, University of Durham, Durham, UK, pp. 55-60, 1992.
- [11] Z. Maalej, "English-Arabic machine translation of nominal compounds," in *Proc. the Workshop on Compound Nouns: Multilingual Aspects of Nominal Composition*, Geneva, 1994.
- [12] A. El-Desouki, A. Abd Elgawwad, and M. Saleh, "A proposed algorithm for English-Arabic machine translation system," in *Proc. the 1st KFUPM Workshop on Information and Computer Sciences (WICS)*, pp. 66-72, 1996.
- [13] H. M. O. Mokhtar, N. M. Darwish, and A. Rafea, "An automated system for English-Arabic translation of scientific text (SEATS)," in *Proc. MT2000: Machine Translation and Multilingual Applications in the New Millennium, the British Computer Society (BCS)*, London, pp. 1-5, 2000.
- [14] Y. Abd Hannouna. (2004). Evaluation of machine translation systems: The translation quality of three Arabic systems. [Online]. Available: <http://aweji.org/index.php/theses-dissertations/154-yasmin-hikmet-abdul-hamid-hannouna>
- [15] B. Hijazi, (2012). Assessment of Google's translation of legal texts. [Online]. Available: <https://www.uop.edu.jo/En/Research/Theses/Documents/Basel%20Abd-Alkareem%20Hijazi.pdf>
- [16] J. Sinclair, *BBC English dictionary*. Haper Collins Publishers, 1992.
- [17] M. Dictionary, *The Macquarie Dictionary*, Sydney: Macquarie Library Pty. Ltd, 1982.
- [18] T. McArthur, *The Oxford Companion to the English Language*, New York: Oxford University Press, 1992.
- [19] D. Summers and P. Stock, *Longman Dictionary of English Language and Culture*, Harlow, England: Longman, 1993.
- [20] M. Ashby, *Oxford Advanced Learner's Dictionary of Current English*, 2000.



Zakaryia Mustafa Slameh Almahasees was born in Jordan on 14/11/1986. Currently, he is a PhD student in machine translation at University of Western Australia, Australia. His main research is translation and technology: machine translation. He is working on English- Arabic machine translation systems. He worked as full-Time lecture for three years at English Department, Najran University, Saudi Arabia. He got his PhD scholarship from Applied Science University, Jordan in 2015. He completed his MA in English language and literature in 2012 language and literature, Jordan and his BA in English in 2008 from Yarmouk University, Jordan.